

Dear Reviewer,

we would like to thank you for thoroughly reading our manuscript. We were pleased to learn that you are generally supporting the publication of the manuscript in ESD. Following your suggestions for improvement, we will increase the level of details regarding the LPJmL simulations. Moreover, we will refine the Results and Discussion Section to improve the balance between different sections.

Please find our point-by-point responses below.

1. Please give more details for the LPJmL simulations you used and discuss their limitations:

Which specific LPJmL version did you use? Our results are based on the modelling output by Lade et al. 2021, who used LPJmL 4.0.002. We will add the version to the manuscript.

Which land-use input did you use? The land-use change implementation used by Lade et al. is based on historical CFT allocation. We will add further details on the land use change implementation in Section 2.1 Estimating the strength of indirect impacts.

Please mention, how you calculate the carbon storage density. I suppose it should be the sum of vegetation, soil, and litter carbon. We use solely vegetation carbon as carbon storage density, in alignment with Lade et al. 2021. We will add this detail to the manuscript.

Please note that using monthly climate inputs (that internally have to be distributed over the days in a month) leads to ambiguity, that can be prevented by using (more up-to-date) daily data e.g., from ISIMIP (GSWP3-W5E5) or TRENDY (CRU-JRA). This is especially relevant, since you compare the differences between runs. I am sure you have checked, that the seed with which monthly values are distributed to days is the same between runs, otherwise you have a climate effect, where you don't expect one. Thank you very much for this remark. Our simulations are based on the modelling output by Lade et al. 2021, who used a fixed seed across runs to distribute monthly values to days.

LPJmL4.X.Y as described by Schaphoff et al. 2018 has 11 PFTs. Your text in (p.6 line 99) could be mistaken so that LPJmL4 would have 12. I believe you use barren land as land cover type 12, but strictly speaking this is not a PFT. Please also add how you calculate it, because you need separate grid cell fractions for CFTs for this ($\text{barren} = 1 - \sum(\text{PFT fpc}) - \sum(\text{CFT fpc})$). Grid cells were classified as barren land whenever the other eleven plant functional types together cover less than 50% of the total grid cell area. We will add this explanation to the manuscript and further clarify that LPJmL only distinguishes eleven PFTs. As described on p.6, l.100 the Earth partition you are referring to is based on the LPJmL run without human land use, so there is no cropland. In particular, $\text{barren} = 1 - \sum(\text{PFT fpc})$.

With LPJmLv4 instead of v5, you are missing N limitation, thus overestimating the CO₂ fertilization effect. Thank you very much for pointing out this potential caveat. We will mention it in Section 3.5 Limitations of the Study.

The sensitivity analysis using 6 more (GCM based) climate input datasets could be integrated better. Currently in the Methods (p. 5 lines 91-92) it is not clear that this is only for sensitivity tests and that the results won't go into the main analysis. We will clarify the fact that the runs driven by the six additional climate models serve as sensitivity analysis only.

2. Details of the clustering approaches.

What exactly are the inputs to the “bottom-up” approach? As far as I understood it, the final effect maps are also going in, so naturally much better clusters are derived than for “top-down”. Is that a fair comparison between the two cluster algorithms?

Like this it appears that the result is clear before hand, and only serves as an argument to focus on the bottom-up clustering results. This is not a problem per se, but could be stated more clearly. I would additionally suggest to shorten the analysis of the “top-down” part, if it is then discarded anyways.

At the same time, the policy implications of using dynamic and thus hard to name and locate clusters could be discussed in more detail.

Yes, it is indeed intuitively clear that the bottom-up clustering approach yields better results with respect to cluster validity indices than the top-down approach. At the same time, as you are mentioning, the resulting clusters are harder to interpret and thereby communicate, which is particularly important if a corresponding study is addressing a non-academic community. Hence, the main message of our manuscript is not the superiority of the bottom-up clustering approach, but a comparison of the two approaches from different perspectives, building awareness for the caveats of both of them. Since this point is apparently not clear in the current version of the manuscript (compare Reviewer 1, Minor Comment 4), we will clarify it both in the Introduction and in the Conclusions.

3. Definition of second-order impacts

Like reviewer R1, I also believe that your current definition of “second-order impacts” based on direct and indirect effects is not sufficiently clear. The example you give rather confuses me, as I would say deforestation IS the vegetation cover change, thus zero-order, and everything that follows next would be direct or first-order impact (like runoff/ET changes).

Maybe the notion of local vs. remote effects can help you to improve the definition? I admit that for remote effects besides river discharge, LPJmL offline is not the right tool.

In order to avoid confusion, we will follow the suggestion of Reviewer 1 and use the more established term “indirect impact” instead of the term “second-order impact” (compare Reviewer 1, Major Comment 2). In particular, we will add the following definition to the first paragraph of the Introduction: “[...] leading to so-called *indirect* impacts. For example, while we consider the decrease in natural vegetation as a *direct* impact of deforestation, the effects of this decrease on carbon storage and water availability are considered *indirect* impacts. Hence, *direct* impacts involve human pressure and one Earth system processes, while *indirect* impacts involve human pressure and two Earth system processes. In particular, we refer to impacts as one-way cause-and effect relations, not including feedback mechanisms.”

4. Justification of using LPJmL offline

Unlike reviewer R1, I believe that it is still worthwhile doing the analysis with LPJmL. Since the authors are using historical land-use and weather patterns, even if not directly coupled, the climate inputs include land-atmosphere interactions to the historical land-use changes.

This is however only true for the factual wLUC simulation, and not the counterfactual noLUC simulation that uses the same climate inputs. This limitation needs to be stated clearly.

As mentioned in the responses to Reviewer 1, Lade et al. deliberately chose an offline approach to be able to isolate particular effects without the need to disentangle overlaying feedback effects (compare Major Comment 1). This scope is not conflicting with the imprint of feedback effects in the historical climate data since the response variable in the estimation of LUC → CC is the modelled carbon storage density, and not atmospheric CO₂.

We will still point the reader to the imprint of feedback effects in Section 2.1 Estimating the strength of indirect impacts.

5. Definition of impacts on climate change

You chose carbon storage density changes as a metric for climate change. Since grid-cells in LPJmL are not the same size, carbon storage density changes in gC/m² does not impact atmospheric CO₂ concentrations in the same way everywhere. Secondly, the translation from carbon storage decline to atmospheric CO₂ increase is certainly not linear. These limitations need to be stated more clearly.

We would like to point out that carbon storage density is only used in the estimation of the effect LUC → CC. The change in natural vegetation cover is not scaled by grid cell area as well and multiplying both and numerator and denominator in the differential quotient would result in the same effect size. Nevertheless, we agree that area is a crucial factor when it comes to the *absolute* direct and indirect impact of human disturbances and corresponding applications of our results should take it into account.

The fact that carbon storage decline translates non-linearly to atmospheric CO₂ increase will be stated in Section 2.1 Estimation the strength of indirect impacts.

Additional comments:

p.2 line 36: If you want to compare with different hierarchical clustering approach, aimed at reducing computational costs please see Ehret et al. 2020 (<https://doi.org/10.5194/hess-24-4389-2020>) or Priesner et al. 2026 (<https://doi.org/10.1140/epjs/s11734-026-02526-1>) who additionally compare different distance metrics. Thank you very much pointing us towards these two articles. Especially the Priesner et al. article fits very well into our argumentation and we will refer to it in the Introduction.

p.4 lines 78-80: The way you define it, you are rather talking about small changes in the land use input, rather than “noise” (numerical non-deterministic behavior?) in the model. Yes, we agree and will update the terminology accordingly.

p.8 lines 161-162: And certainly also the different methodology, by regressing values for each year instead of a 30-year average, right? Yes, indeed. We will reformulate the corresponding sentence in l.164 to clarify that it refers to the phenomenon of the patchy pattern as well.

p.10 lines 191-192: Can you really draw this conclusion from the Figure (and what does “best explain” mean – “highest RS”, or per no. of clusters so “most top-left”)? It is hard to see (you also highlighted “non climate” partitions), maybe plot only 3 colors for different symbols, red for climate and blue for PFT, and orange for biome to confirm this. By “best explained/represented” we mean that the corresponding partitions stand out with respect to their RS index compared to partitions of similar size (number of clusters). As being explained in the caption of Figure 6, we marked all partitions that fulfil this criterium. While there is a clear dominance of e.g., climate-based partitions in the case CC → RO, the dominating classification criterion is not as clear in other cases. Since the proposed coloring scheme would be difficult due to partitions combining classification criteria, we highlighted the dominating classification criterion in bold letters. We will clarify both the procedure and interpretation in the text.

p.13 lines 266-267: I don't find this very surprising, as each PFTs parameters describe a direct relationship with climate, especially in the temperature limits and water/radiation requirements. It would be interesting to single out the climate from the CO₂ effect, by increasing CO₂ regularly but keeping the climate as 1901-1930 and vice versa. We assume you are referring to lines 263-265. While we agree that the effect of climate change on vegetation cover is a

very direct one, we do not see how this explains the fact that the bottom-up RS index perform best for the effect LUC → RO, while the bottom-up BR index performs best for the effect CC → LUC. We agree that isolating the effect of changing CO₂ values could yield interesting insights, as studies like Piao et al. 2007 demonstrated.

Fig. 6: Some combinations are hard to distinguish, e.g. same symbol and similar color as in “climates P” and “continents biomes”. In panel B the names are offset from the boxes (which I believe) they belong to. There is no clear method for when to draw boxes, right? It appears a bit arbitrary. In general everything is very small, even when zooming. Yes, in Figure 6B, the names are offset from the boxes in order to increase readability. We will try to find an allocation which is visually clearer. The boxes were drawn to highlight partitions, which stand out in RS compared to partitions of similar size. We will further clarify this aspect in the manuscript. Overall, we will replot the figure in order to avoid same symbol/similar colour- combinations and check if increasing symbol size etc. increases readability of the plot.

Fig. A8: Labels appear erroneous. “Boreal broadleaved evergreen” does not exist in LPJmL. “Temperate needleleaved evergreen” is missing. Thank you very much for pointing out this mistake. “Boreal broadleaved evergreen” should be “boreal broadleaved summergreen” and “temperate needleleaved” should be “temperate needleleaved evergreen”. We will correct this mistake both in the supplementary code and in the figure.

Fig. D1A: The dashed boxes are off. We will correct this figure,